

ABSTRAK

Perkembangan platform e-commerce menyebabkan jumlah produk yang semakin banyak sehingga pengguna mengalami *information overload* dalam menemukan produk yang sesuai. Salah satu solusi yang digunakan adalah sistem rekomendasi berbasis *Content-Based Filtering* (CBF), namun metode ini masih menghadapi permasalahan *popularity bias*, yaitu kecenderungan merekomendasikan produk populer dibandingkan produk *long-tail*. Penelitian ini bertujuan menganalisis tingkat *popularity bias* pada *Pure Content-Based Filtering* serta mengevaluasi efektivitas *Hybrid Content-Based Filtering* dengan mekanisme *Inverse Popularity Weighting*. Penelitian menggunakan dataset Amazon Reviews 2023 kategori *Toys and Games*. Representasi produk dibangun menggunakan TF-IDF dan *Cosine Similarity*, sedangkan evaluasi dilakukan secara *offline* menggunakan protokol *last-out* dengan metrik akurasi dan *beyond-accuracy*. Hasil penelitian menunjukkan bahwa model *Hybrid Content-Based Filtering* mampu meningkatkan *Long-tail Coverage@10* sebesar 101,5% dan menurunkan *Average Popularity@10* sebesar 82,8% dibandingkan *Pure Content-Based Filtering*, meskipun terjadi penurunan pada metrik akurasi. Selain itu, model berhasil diimplementasikan dalam simulasi sistem rekomendasi berbasis web sehingga menunjukkan potensi penerapan pada platform e-commerce.

Kata Kunci: Sistem Rekomendasi, *Content-Based Filtering*, *Popularity Bias*, *Long-tail*, TF-IDF.

ABSTRACT

The rapid growth of e-commerce platforms has increased the number of available products, making it more difficult for users to find relevant items. Content-Based Filtering (CBF) is widely used to address this problem but still suffers from popularity bias, where popular products are recommended more frequently than long-tail products. This study aims to analyze the popularity bias of Pure Content-Based Filtering and evaluate a Hybrid Content-Based Filtering approach incorporating Inverse Popularity Weighting. The study uses the Amazon Reviews 2023 – Toys and Games dataset. Product representations are generated using TF-IDF and Cosine Similarity, while the models are evaluated offline using the last-out protocol with both accuracy and beyond-accuracy metrics. The results show that the proposed Hybrid Content-Based Filtering improves Long-tail Coverage@10 by 101.5% and reduces Average Popularity@10 by 82.8% compared with the Pure Content-Based Filtering baseline, although a decrease in accuracy is observed. The proposed model has also been implemented in a web-based recommendation system simulation, demonstrating its potential application in e-commerce platforms.

Keywords: Recommender System, Content-Based Filtering, Popularity Bias, Long-Tail, TF-IDF.